

# SCIENTIFIC REPORTS

OPEN

## The utility of Next Generation Sequencing for molecular diagnostics in Rett syndrome

Silvia Vidal<sup>1</sup>, Núria Brandi<sup>2</sup>, Paola Pacheco<sup>1</sup>, Edgar Gerotina<sup>1</sup>, Laura Blasco<sup>1</sup>, Jean-Rémi Trotta<sup>3</sup>, Sophia Derdak<sup>3</sup>, Maria del Mar O'Callaghan<sup>4,5,6</sup>, Àngels Garcia-Cazorla<sup>4,5,6</sup>, Mercè Pineda<sup>5</sup>, Judith Armstrong<sup>1,5,6</sup> & Rett Working Group\*

Received: 23 June 2017

Accepted: 29 August 2017

Published online: 25 September 2017

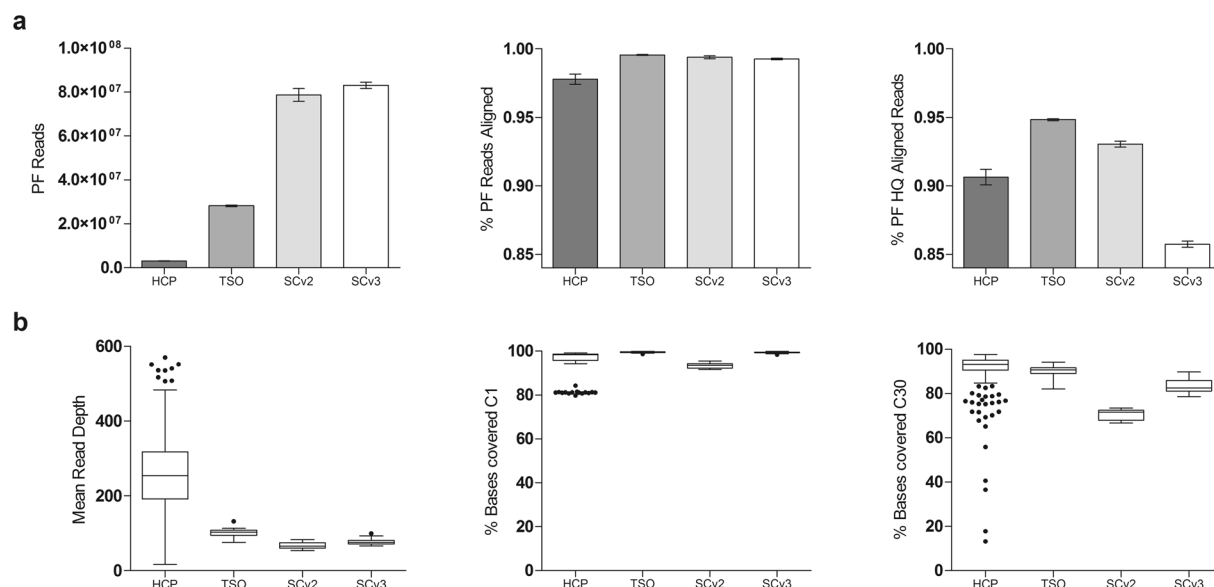
Rett syndrome (RTT) is an early-onset neurodevelopmental disorder that almost exclusively affects girls and is totally disabling. Three genes have been identified that cause RTT: *MECP2*, *CDKL5* and *FOXG1*. However, the etiology of some of RTT patients still remains unknown. Recently, next generation sequencing (NGS) has promoted genetic diagnoses because of the quickness and affordability of the method. To evaluate the usefulness of NGS in genetic diagnosis, we present the genetic study of RTT-like patients using different techniques based on this technology. We studied 1577 patients with RTT-like clinical diagnoses and reviewed patients who were previously studied and thought to have RTT genes by Sanger sequencing. Genetically, 477 of 1577 patients with a RTT-like suspicion have been diagnosed. Positive results were found in 30% by Sanger sequencing, 23% with a custom panel, 24% with a commercial panel and 32% with whole exome sequencing. A genetic study using NGS allows the study of a larger number of genes associated with RTT-like symptoms simultaneously, providing genetic study of a wider group of patients as well as significantly reducing the response time and cost of the study.

Rett syndrome (RTT; MIM# 312750) is a neurodevelopmental disorder of early onset that affects girls almost exclusively. RTT was originally described in the 1960s by Andreas Rett<sup>1</sup>. This syndrome is first recognized in infancy with a period of apparently normal development (up to the age of 6–18 months), followed by a regression characterized by loss of speech and purposeful hand use and motor apraxia that may be associated with epilepsy and dysautonomic features, including disturbed breathing, sleep and gastrointestinal motility<sup>1,2</sup>. RTT has an incidence of 1:10,000 live female births and is the second cause of intellectual disability after Down's syndrome in females<sup>3</sup>. RTT was clinically cataloged into classic and atypical forms of the disease. However, these criteria have undergone several updates over the past three decades. Consensus criteria have been established that distinguish RTT patients into the individual classifications of classic or typical RTT and the atypical or variant forms of RTT<sup>4–6</sup>.

A large number of reports support the evidence that mutations in the Methyl CpG binding protein 2 gene (*MECP2*; MIM \*300005) are the major causes of classical RTT<sup>7,8</sup>. Over 95% of the cases are explained by more than 800 reported mutations in the methyl CpG-binding protein 2 gene (*MECP2*) (RettBASE: *MECP2* Variation Database)<sup>9,10</sup>. *MECP2* is a transcriptional regulatory protein, and in its absence, a large number of genes exhibit abnormal expression with implications in the balance between synaptic excitation and inhibition<sup>11</sup>.

Although the majority of RTT patients have mutations in the *MECP2* gene<sup>10</sup>, approximately 5% of classical RTT and 25% of variant RTT patients are negative for *MECP2* mutation<sup>6,12</sup>. In this variant RTT group of patients, some have mutations in other genes that are also associated with RTT: cyclin-dependent kinase-like5 (*CDKL5*; MIM \*300203), which is described in individuals with an early seizure onset variant of RTT<sup>13</sup>, and Forkhead box protein G1 (*FOXG1*; MIM \*164874), which is responsible for the congenital variant of RTT<sup>14</sup>. In addition, with

<sup>1</sup>Molecular and Genetics Medicine Section, Hospital Sant Joan de Déu, Barcelona, Spain. <sup>2</sup>Facultat de Medicina, Universitat de Barcelona, Barcelona, Spain. <sup>3</sup>Centro Nacional de Análisis Genómica (CNAG-CRG), Center for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. <sup>4</sup>Neurology Service, Hospital Sant Joan de Déu, Barcelona, Spain. <sup>5</sup>Institut de Recerca Pediàtrica Hospital Sant Joan de Déu, Barcelona, Spain. <sup>6</sup>CIBER-ER (Biomedical Network Research Center for Rare Diseases), Instituto de Salud Carlos III, Madrid, Spain. \*A comprehensive list of consortium members appears at the end of the paper. Correspondence and requests for materials should be addressed to J.A. (email: [jarmstrong@sjdhospitalbarcelona.org](mailto:jarmstrong@sjdhospitalbarcelona.org))



**Figure 1.** Comparison of main coverage metrics. Average of all samples analyzed for the four different approaches. **(a)** Bar plots with 95% confidence interval for the four approaches. Alignment metrics: passing filter (PF) reads; percentage of PF reads that aligned to the reference sequence; and percentage of PF reads that were aligned to the reference sequence with a mapping quality of Q20 or higher signifying. **(b)** Tukey boxplots: Mean read depth; percentage of bases covered at C1; and percentage of bases covered at C30.

the introduction of next generation sequencing (NGS), other genes without previous relation to RTT have been associated with RTT-like phenotypes, such as myocyte-specific enhancer factor 2C (*MEF2C*; MIM \*600662) and transcription factor 4 (*TCF4*; \*602272)<sup>15</sup>. However, the etiology of a subset of patients with a clinical diagnosis of RTT or RTT-like symptoms remains unknown.

In recent years, NGS has emerged as a potentially powerful tool for the study of this type of genetic disease<sup>16,17</sup>. Now, multiple genes can be sequenced at the same time and at a comparable cost to the Sanger analysis of only one single gene<sup>18,19</sup>. However, for genetic diagnostics, Sanger sequencing still remains necessary to validate the detected variants to avoid false positives.

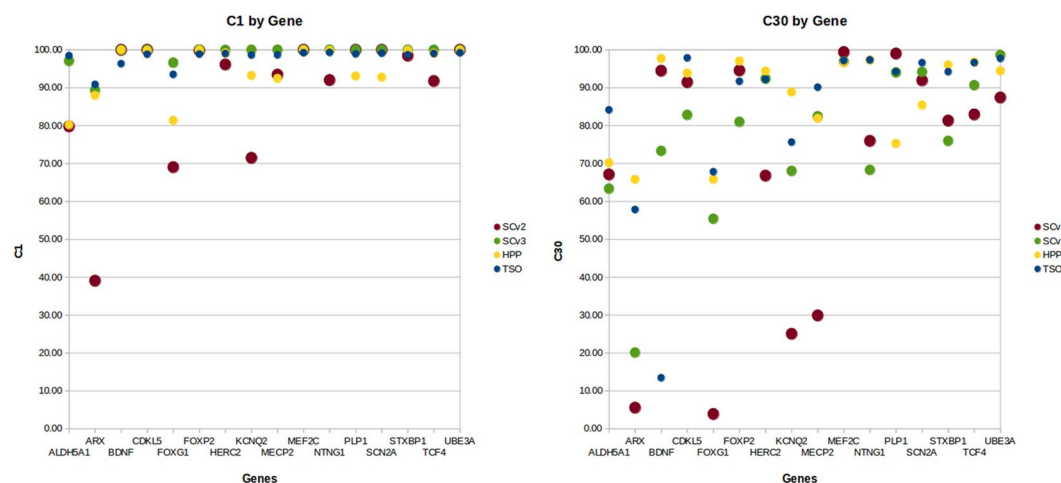
Here, we present the retrospective results from our group using Sanger sequencing and three different approaches for library construction of NGS libraries. Our main goal was to assess the relative advantages and disadvantages of each methodology for diagnostic purposes.

## Results

A total of 1577 patients with RTT-like syndrome were genetically analyzed between 1999 and 2016 at Sant Joan de Déu Hospital, Barcelona (see Fig. S1). These patients had been diagnosed following the usual clinical parameters<sup>6</sup> and according to the recently revised RTT Search International Consortium criteria and nomenclature<sup>6</sup>. Throughout the Sanger period, 84% of RTT classical phenotypes and 16% of atypical phenotypes were recruited for study. From 2012, a wider group of patients was recruited, not only RTT classical and atypical but also with RTT-like features, due to the incorporation of the NGS technologies.

Four approaches were used to genetically analyze all RTT patients who were recruited. A total of 1341 patients were studied based on Sanger sequencing (SS). The high throughput approaches were based on three distinct gene library preparations. First, 242 patients were studied using the Haloplex Custom Panel (HCP, Agilent Technologies) (Santa Clara, California), including 46 that came from a negative study of SS. Second, 51 patients were studied using the TruSight One panel (TSO, Illumina) (San Diego, California), including 11 who came from a negative study of HCP. Finally, 22 patients declared as negative-SS and 3 declared as negative-HCP were studied using the SeqCap EZ Human Exome v2.0 and v3.0 (WES, Roche NimbleGen) (Madison, Wisconsin). SS was used to study the exons and surroundings of the three genes associated with RTT (*MECP2*, *CDKL5* and *FOXG1*). HCP is a custom panel of 17 genes designed to cover the exons and surroundings of genes associated with RTT-like phenotypes. TSO targets 4,813 genes associated with a clinical phenotype, including RTT-like phenotypes. The WES kit was used to target all human coding exons, including genes covered by panels.

**Run and Mapping Quality.** A summary of quality control data is represented in Fig. 1a. The total number of passing filter reads (PF reads) was approximately 9 times higher in TSO (approx. 28 million reads) than HCP (approx. 3 million reads), whereas WES (approx. 80 million reads) was 3 times higher than TSO. The percentage of unique-mapping reads aligned to the reference sequence was higher in TSO (99.5%) and WES (99.4% and 99.2%) than HCP (97.8%). In addition, the unique-mapping reads aligned with a mapping quality Q20 or higher, indicating that the aligner estimates a 1/100 (or smaller) chance of a wrong alignment, was also higher in TSO (94.8%) and WES (93.0 and 85.7%) than HCP (90.6%).



**Figure 2.** Comparison of coverage in 17 RTT-like genes. Scatter plots of average of the coverage at C1 and C30 of all samples analyzed for the four different approaches.

**Target Regions, Read Depth and Coverage.** The diagnosis regions of interest for a given set of genes that are related to RTT-like features is defined as the sum of the targets defined by the coding bases of the exons plus a 25 bp flanking region. Three sets of genes of interest were defined where each set included the previous one: RTT list, which included the 3 genes related to RTT; RTT-like list, 3+60 genes related to RTT-like disease; and RTT+EEP list, 3+14 RTT-like genes and 526 genes related to RTT-like features and epileptic encephalopathy (EEP) (Table 1S). Depending on the NGS methods and analyses performed, we evaluated the set of 17 genes common to both panels (TSO and HCP) and the WES, or the set of 605 genes most included in the TSO and all of them in the WES. The gene lists are shown in Supplementary Table 1S.

The performance of the three approaches was compared as if they were four since the WES samples were captured with two different kits, i.e., with SureSelect version 2.0 in 2011 (SCv2) and version 3.0 (SCv3) in 2014. In this study, 30 reads per base was considered the minimum coverage (C30) for high-sensitivity heterozygote detection. For the RTT-like list, average mean read depths of  $262\times$ ,  $99\times$ ,  $67\times$  and  $77\times$  were obtained by HCP, TSO, SCv2 and SCv3, respectively. Ninety percent of the targeted bases were covered at C30 by HCP, and for TSO, 88% of the bases were, while WES was 70% in 2011 and 84% in 2014. These mean coverage results are further compared in Fig. 1b, displaying variation among samples with the four methods referring to the mean read depth and bases covering C1 and C30. Both the coverage and the uniformity of the capture sequencing were better in TSO and WES than HCP. Deepening the coverage data, we analyzed the C1 and C30 of the 16 genes related to an RTT-like phenotype included in HCP considering the four approaches. For the C1 and C30 plots (Fig. 2), we used the raw CCDS coding exon coordinates as common references to compare as fairly as possible the different capture approaches. In this analysis, we discarded the *SHANK3* gene since this gene does not have a CCDS ID (Fig. 2).

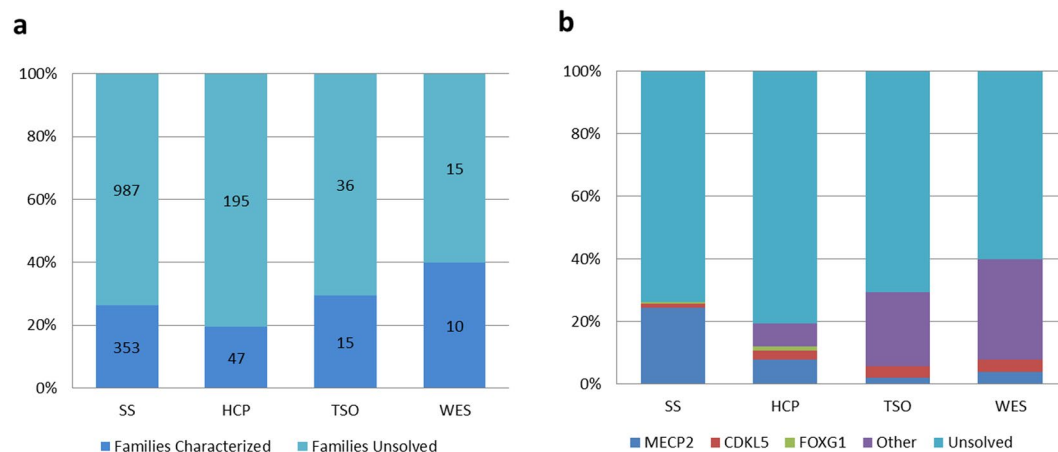
**Variant detection.** According to diagnostic quality standards, all regions not reaching the required  $30\times$  must be Sanger sequenced; from HCP samples, 2 regions of the *MECP2* gene were sequenced by SS. Non-targets were sequenced by SS in TSO and WES. To identify the potential mutations, we checked the variants by matching their affected phenotypes and inheritance patterns of respective genes checked by SS of the index cases and their progenitors. Moreover, we considered the pathogenicity predictors (Mutation Tester, Polyphen-2 and Sorts Intolerant From Tolerant) and reviewed the literature, the RettBASE: MECP2 Variation Database, the Exome Aggregation Consortium (ExAC), HGMD® Professional 2016.4 and The Database of Short Genetic Variation (dbSNP). The potentially pathogenic variants detected in the genes that were not *MECP2*, *CDKL5* or *FOXG1* are shown in Table 1. All results are shown in Supplementary Table S2.

**Sanger Sequencing-SS.** A total of 1341 patients were studied by SS between 1999 and 2012, and 375 (22 by MLPA) were genetically diagnosed (28%) with mutations in *MECP2*, *CDKL5* or *FOXG1*. During this period of time, the workflow was to study the four exons of the *MECP2* gene to detect SNVs or short indels and MLPA for gross rearrangements. Excluding large rearrangements detected by MLPA, a total of 293 patients with RTT classic and 36 with atypical phenotypes had mutations in *MECP2*. Then, the *CDKL5* and *FOXG1* genes were studied in the patients without mutations in *MECP2*. No patient diagnosed with RTT-classic had mutations in *CDKL5* or *FOXG1*, and 15 patients with atypical RTT had mutations in *CDKL5* and 9 in the *FOXG1* gene.

**Haloplex Custom Panel - HCP.** A total of 242 patients were studied with HCP between 2012 and 2016, and 53 (6 by MLPA) patients were genetically diagnosed (22%). Excluding large rearrangements detected by MLPA, 29 patients had mutations in RTT genes and 90% in *MECP2*. From 18 patients with mutations in RTT-like genes, it is remarkable that the majority of these mutations were in the *STXBP1* gene, which is associated with early infantile epileptic encephalopathy (EIEE4)<sup>20</sup>, and the *TCF4* gene, which is associated with Pitt-Hopkins syndrome<sup>21</sup>.

Num. Patients	Gene	OMIM number	Transcript	Type of seq. Change	cDNAchange	Proteinchange	dbSNP	Mutation taster, SIFT, PROVEAN, PolyPhen-2 scores
<b>Potentially pathogenic mutation detected by HCP</b>								
2	STXBP1	602926	NM_003165	Missense	c.874C>T	p.Arg292Cys	—	Disease causing, 0, 0.996, −7.58
2	STXBP1	602926	NM_003165	Missense	c.875G>A	p.Arg292His	rs796053361	Disease causing, 0, 1, −4.74
1	KCNQ2	602235	NM_172107	Missense	c.593G>A	p.Arg198Gln	rs796052621	Disease causing, 0, 1, −3.58
1	KCNQ2	602235	NM_172107	Missense	c.637C>T	p.Arg213Trp	rs118192203	Disease causing, 0, 1, −7.19
1	SLC2A1	138140	NM_006516	Missense	c.805C>T	p.Arg269Cys	rs200247956	Disease causing, 0, 1, −7.79
1	STXBP1	602926	NM_003165	Missense	c.1216C>T	p.Arg406Cys	rs796053367	Disease causing, 0, 1, −7.86
1	STXBP1	602926	NM_003165	In-frame deletion	c.124_126delTCC	p.Ser42del	—	Disease causing, NA, NA, −10.71
1	STXBP1	602926	NM_003165	Splicing variant	c.326-3C>G	Miss-splicing	—	NA, NA, NA, NA
1	STXBP1	602926	NM_003165	Missense	c.704G>A	p.Arg235Gln	—	Disease causing, 0, 1, −3.79
1	TCF4	602272	NM_001243236	In-frame indel	c.1169_1175delTAGAAAGinsAAA	p.Leu390Ter	—	Disease causing, NA, NA, NA
1	TCF4	602272	NM_001243236	Missense	c.1733G>A	p.Arg578His	rs121909123	Disease causing, 0, 1, −4.73
1	TCF4	602272	NM_001243236	Nonsense	c.1774C>T	p.Gln592Ter	—	Disease causing, NA, NA, NA
1	TCF4	602272	NM_001243236	Frameshift deletion	c.514_517delAAAG	p.Lys172PhefsTer61	rs398123561	Disease causing, NA, NA, NA
<b>Potentially pathogenic mutation detected by TSO</b>								
1	MEF2C	600662	NM_001193347	Missense	c.48C>G	p.Asn16Lys	—	Disease causing, 0.013, 0.995, −5.35
1	MEF2C	600662	NM_001193347	Frameshift deletion	c.989_990delGT	p.Gly330AspfsTer7	—	Disease causing, NA, NA, NA
1	SCN2A	182390	NM_001040142	Missense	c.3631G>A	p.Glu1211Lys	rs387906684	Disease causing, 0, 0.995, −3.82
1	SCN2A	182390	NM_001040142	Missense	c.5317G>A	p.Ala1773Thr	—	Disease causing, 0, 1, −3.68
1	SYNGAP1	603384	NM_006772	Frameshift deletion	c.2019delA	p.Thr674ProfsTer36	—	Disease causing, NA, NA, NA
1	SYNGAP1	603384	NM_006772	Frameshift deletion	c.1782delC	p.Leu595CysfsTer55	rs587780470	Disease causing, NA, NA, NA
<b>Potentially pathogenic mutation detected by WES</b>								
1	CACNA1I	608230	NM_021096	Missense	c.4435C>T	p.Leu1479Phe	—	Disease causing, 0.397, 0.756, −1.53
1	CHRNA5	118505	NM_000745	Missense	c.748C>A	p.Pro250Thr	—	Disease causing, 0.301, 1, −6.07
1	GABBR2	607340	NM_005458	Missense	c.1699G>A	p.Ala567Thr	—	Disease causing, 0.002, 0.999, −3.48
1	GRIN2B	138252	NM_000834	Missense	c.1657C>A	p.Pro553Thr	—	Disease causing, 0.001, 0.975, −6.8
1	HCN1	602780	NM_021072	Missense	c.1159G>T	p.Ala387Ser	—	Disease causing, 0.002, 0.767, −2.76

**Table 1.** Potentially pathogenic and causative SNVs detected, excluding RTT genes. Variant effect predictors web tools used: Mutation taster (<http://www.mutationtaster.org/>); SIFT-PROVEAN, SIFT scores ranged from 0–1, where 0 is predicted to be most damaging and Protein Variation Effect Analyzer (PROVEAN) score  $\leq -2.5$ , the protein variant is predicted to have a “deleterious” effect, while if the PROVEAN score is  $> -2.5$ , the variant is predicted to have a “neutral” effect ([http://provean.jcvi.org/genome\\_submit\\_2.php](http://provean.jcvi.org/genome_submit_2.php)); and Polyphen-2, ranged from 0–1, where 1 is most likely to be damaging (<http://genetics.bwh.harvard.edu/pph2/>).



**Figure 3.** Statistics results for genetic diagnosis. Each columns represents different detection methodology used (SS, HCP, TSO and WES) (a) Percentage and total of families characterized and unsolved by the four different approaches. (b) Percentage of patients with *MECP2*, *CDKL5*, *FOXG1*, other genes with pathogenic mutations and the unsolved cases for the Sanger Sequencing (SS) and the three NGS approaches. Total of mutations found are detailed in Table S2.

**TruSight One panel - TSO.** Fifty-one patients were studied by TSO since 2015, and 15 patients were genetically diagnosed (29%). Three patients had mutations in RTT genes and 12 in RTT-like genes. In these 12 patients, we detected 2 SNVs in *MEF2C* and 2 in *SCN2A* genes, which are associated with mental retardation, stereotypic movements, epilepsy, and/or cerebral malformations<sup>22–24</sup>; 2 SNVs in the *SYNGAP1* gene, which is associated with mental retardation<sup>24,25</sup>; a deletion of *IQSEC2* and *KDM5C* genes, which are associated with mental retardation<sup>26</sup>; a gross deletion in chr15 (chr15:22,833,395–28,567,298), which is associated with Prader-Willi syndrome<sup>27</sup>; and a duplication in chr14 (14q32.11–q32.33(90949120–107287505)), which is associated with mental retardation and development delay<sup>28</sup>. All CNVs detected through the read depth and detected by NGS were confirmed by CGHarray.

**Whole Exome Sequencing - WES.** Twenty-five patients were studied by WES in 2011 (SCv2) and 2014 (SCv3). Five occurred in genes previously associated with neurodevelopmental disorders with features similar to those of RTT syndrome. We identified four mutations in genes such as *HCN1*<sup>29</sup> and *GRIN2B*<sup>30</sup>, which are associated with early infantile epileptic encephalopathy; *SLC6A1*, which is associated with epilepsy and myoclonic-atic seizures<sup>31</sup>; *TCF4*, which is associated with Pitt–Hopkins syndrome<sup>21</sup>; and *SCN1A*, which is associated with Dravet syndrome<sup>32</sup>. The other potentially pathogenic variants that were identified occurred in genes that had not been linked to any genetic disorder<sup>33</sup>. However, there was an enrichment of genes with a potential role in neuronal biology and functionality, such as the gamma-aminobutyric type B receptor subunit 2 (*GABBR2*), the neuronal acetylcholine receptor subunit alpha-5 (*CHRNA5*), and the neuronal voltage-gated calcium channel (*CACNA1I*).

**Genetic diagnosis.** The total of cases and statistical results of genetic diagnoses of the RTT-like patients are listed in Fig. 3a. We identified the genetic cause for 353/1341 patients studied with SS. All characterized patients (26.3%) had mutations in RTT genes (*MECP2*, 24.5%; *CDKL5*, 1.1%; *FOXG1*, 0.7%). Forty-seven of 242 patients (19.4%) were genetically diagnosed by HCP; for these patients, 7.9% had mutations in the *MECP2* gene, 2.9% in *CDKL5*, 1.2% in *FOXG1* and 7.4% in other genes related to RTT-like phenotypes. Fifteen of 51 patients (29.4%) were genetically diagnosed by TSO; for these patients, 2.0% had mutations in the *MECP2* gene, 3.9% in *CDKL5* and 23.5% in other genes related to RTT-like phenotypes. Ten of 25 patients (40.0%) were genetically diagnosed by TSO; for these patients, 4.0% had mutations in the *MECP2* gene, 4.0% in *CDKL5* and 32.0% in other genes related to RTT-like phenotypes (Fig. 3b).

In addition of these results, it has been identified by MLPA twenty four patients with gross deletions which are not included in the statistical results. Twenty one patients with gross deletions in *MECP2*, where one have a deletion of exon 1 and 2, six patients in exon 3 and fourteen patients in exon 3 and 4. Two patients have a deletion in *CDKL5*, one with a deletion in exon 1 and 2 and other in exon 8. And three patients with a deletion *FOXG1*.

## Discussion

Over the last few years, several genes have been associated with RTT-like phenotypes<sup>17,33–37</sup> due to the incorporation of NGS. Traditional detection methods for individual genes such as SS can only provide a limited mutation spectrum of the disease, consuming a great amount of time. NGS, which has revolutionized molecular genetics research, is a high-throughput method capable of rapidly sequencing a large number of genes in parallel and providing large datasets<sup>33–38</sup>. This study reports a comparison among different sequencing methods used across the last decade. We first developed a custom panel for molecular diagnosis of RTT and RTT-like phenotypes covering 17 genes related to RTT-like disease and then used a commercial panel that includes these genes and others also



associated with the phenotype. Furthermore, we performed WES to identify unsolved families via SS or a custom panel.

High performance standards are essential in a clinical diagnostic setting. Even though the mean read depth in HCP is ostensibly higher than other methods, with regions covering more than  $500\times$ , the uniformity of the capture is uneven. Figure 1b shows that some samples from different captures are in the border or below the quality standards of diagnosis. In spite of the fact that TSO has a slightly lower coverage at C30 than HCP, the homogeneity and uniformity of all the samples analyzed is more accurate. Even though the number of target region bases not reaching C30 is much higher in WES than in panel sequencing, it should be noted that the quality of the analysis improved in these few years (2011 to 2014). Moreover, these coverage comparatives could be inequitable, and we have to take into account that these WES are not performed with very high coverage and we cannot eliminate duplicates of PCR of the HCP. Thus, we not only compared the C30 but also the C1, which is the region that could potentially be covered if adequate coverage was made. Comparing the C1, TSO and SCv3 could potentially be better coverage than HCP and SCv2<sup>39</sup>.

A large cohort of Spanish patients who exhibited clinical features associated with RTT or RTT-like phenotypes was recruited over three decades. Before the implementation of NGS, the cohort of patients was less clinically homogenous. Today, the cohort presents an increase in the number of patients who were RTT-like compared to classic or atypical RTT. This is likely due to the limitation of the technique because it takes time to analyze the 3 genes associated with RTT using SS and clinicians only send samples from patients who have fulfilled the clinical criteria as RTT.

Regarding additional information of this study, with a more extensive panel, the rate of the families characterized was higher. We characterized 40% of the families analyzed by WES and only 26% by SS. It is obviously less costly to sequence a panel of patients than an exome per trio (patient and progenitors) due to the much smaller capture region. Moreover, the development of panels targeting most clinically relevant genes, such as TSO or HCP, would be more cost-effective than standard WES. However, it must be taken into consideration that our HCP panel required a cost for development and validation, and TSO and WES are both labelled for research use only. For these reasons, proper validation by the laboratory is required before implementation in patient diagnoses. On the other hand, data obtained in a HCP or TSO are easier to interpret than WES results and also the computational requirements for a panel of a few genes is lower than a WES<sup>40</sup>.

Regarding the NGS results, the *MECP2* gene remains the major mutated gene in our cohort (82% of all positive results), followed by *CDKL5* (6%) and *FOXG1* (3%) genes. One of the important findings is that two genes related to other phenotypes were found to be more frequently mutated than in other RTT-like phenotypes. Eight patients had pathogenic mutations in the *STXBP1* gene. The clinical features in patients with a mutation in this gene, such as developmental regression in the neonatal period or infancy, hypotonia, poor visual pursuit, seizures and epileptic encephalopathy, could fit perfectly with the RTT phenotype. Five patients had pathogenic mutations in the *TCF4* gene and one in the *UBE3A* gene, which are associated with Pitt-Hopkins syndrome and Angelman syndrome, respectively. Mutations in these genes cause severe neurologic features such as poor or absent speech development, delayed motor development, seizures, and hypotonia with an onset during the first year of life. Moreover, the patients with Pitt-Hopkins have morphologic characteristics, such as deep-set eyes and fleshy ears, but in our cohort, we found patients without any of these features. We also found 2 patients with mutations in the *MEF2C* gene (mental retardation, stereotypic movements, epilepsy, and/or cerebral malformations) and 2 patients with mutations in the *SYNGAP1* gene (mental retardation, autosomal dominant 5). The fact that there are the same neurologic features commented on previously highlights that *SYNGAP1* has behavioral psychiatric manifestations identical to RTT, that is, autism features and regression of motor development.

It is also remarkable that, in addition to *STXBP1* (EEIE4), we found mutations in genes related to EEIE: *KCNQ2* (EEIE7), *SCN1A* (EEIE6), *SCN2A* (EEIE11), *GRIN2B* (EEIE27) and *HCN1* (EEIE24) genes. Furthermore, we also found members of the solute carrier families related to epilepsy: *SLC6A1* (myoclonic-atonic epilepsy) and *SLC2A1* (idiopathic generalized epilepsy) genes<sup>33,35</sup>.

Finally, mutations in genes without related phenotypes were detected by WES: *GABBR2*, *CHRNA5* and *CACNA1I*. Even though these genes are quite unknown, they have a potential role in neuronal biology and functionality. Further functional studies are required to consider these as the genetic cause of the disease. The clinical characteristics and detected variants of these patients are summarized in Lucariello *et al.*<sup>33</sup>.

In this study, we focused on the comparison of a single nucleotide variant (SNV) and insertion/deletion detection, since these approaches are highly sensitive and specific to detect these types of mutations. Therefore, our diagnostics workflow includes MLPA (Multiplex Ligation-dependent Probe Amplification) of the *MECP2*, *CDKL5*, *FOXG1* and *TCF4* genes. aCGH (array comparative genomic hybridization) can be used for patients who have to be analyzed by WES. Although copy number variations (CNV) are difficult to detect with these NGS methods, we performed a preliminary study with TSO read depths of patients without detecting SNVs, trying to detect CNVs via bioinformatics methods. We were able to detect 3 *de novo* pathogenic CNVs corroborated later by aCGH: two deletions, one of the *IQSEC2* gene and one of the *KDM5C* gene, and another in chr15:22,833,395-28,567,298. One gross duplication in the long arm of chr14:90,949,120-107,287,505 was found.

In summary, the genetic study by NGS allows study of a larger number of genes associated with RTT-like symptoms simultaneously, allowing genetic study of a wider group of patients<sup>10,41</sup>. These detected variants identified by NGS may modify the initial clinical diagnosis to other neurodevelopmental syndromes, or determine new candidate genes related to RTT-like symptoms, providing the clinician with more information and clues that could help in the prevention of future symptoms or in the pharmacologic therapy. For instance, the use of D-serine as a dietary supplement for the enhancement of glutamatergic neurotransmission and/or excitatory/inhibitory neurotransmitter imbalance affected in patients with mutations in N-Methyl-D-aspartate receptors such as *GRIN2B* gene<sup>42</sup>. We could conclude that TSO has the best cost-efficiency of all technologies used and could offer timely responses for clinical diagnosis. However, performing a WES in families to characterize the

family and identify new candidate genes should be included in the target panels. In addition, study of the progenitors remains essential for their characterization as well as the need for functional studies in newly discovered genetic variants.

## Material and Methods

**Patients and DNA extraction.** A cohort of 1577 Spanish patients who exhibited clinical features associated with RTT or RTT-like phenotypes was recruited at Sant Joan de Déu Hospital in Barcelona, Spain, from different Spanish Hospitals. Genomic DNA samples were extracted from peripheral blood leukocytes using the Puregene DNA Isolation kit (Gentra System, Minneapolis, USA) following the manufacturer's instructions. All DNA samples were quantified using a Qubit 2.0 Fluorometer (Invitrogen), and the DNA purity was quantified by calculating the absorbance ratio (A260/280) with a NanoDrop 1000 (Thermo Fisher).

**Ethical issues.** The study was approved by the ethical committees of Hospital Sant Joan de Déu, CEIC: Comitè d'Ètica d'Investigació Clínica- Fundació Sant Joan de Déu (internal code: PIC-101-15). Patients or their parents signed informed consent for genetic studies, and tissue samples from patients and controls were obtained according to the Helsinki Declaration of 1964, as revised in 2001.

**Molecular analysis.** A total of 1341 patients were sequenced by SS for the 3 RTT genes (*MECP2*, *CDKL5* and *FOXG1*); 242 patients by HCP; 51 patients by TSO; and 24 patients with their healthy progenitors by WES. All patients analyzed by WES had an aCGH performed first with a normal or inconclusive profile.

**Copy Number Variations - CNVs.** *Multiplex Ligation-dependent Probe Amplification (MLPA).* *MECP2*-MLPA was performed using a SALSA kit P015, *CDKL5*-MLPA with a SALSA kit P189 and *FOXG1-TCF4*-MLPA with a SALSA kit P075 (MRC-Holland, Amsterdam, The Netherlands) according to the manufacturer's instructions.

*Array Comparative Genomic Hybridization (aCGH).* The aCGH analysis was performed at Bioarray (Genetic diagnosis, Alicante, Spain) using two different platforms: human genome Cytoarray Plus 180K and 400K (Agilent Technologies, Santa Clara, CA, USA). All genomic coordinates are in build GRCh37/hg19.

**NGS: Library Preparation.** Library preparation was conducted according to the manufacturers' instructions. HCP and TSO panels were created at Sant Joan de Déu Hospital and WES at the National Center for Genomic Analysis (CNAG).

*HaloPlex Custom Panel (HCP).* We designed a custom-made panel with 17 genes associated with a RTT-like phenotype based on the evidence curated in the Online Mendelian Inheritance in Man (OMIM). The genes are shown in Supplementary Table S3. Amplicon libraries were prepared using the Agilent HaloPlex Target enrichment system, for Illumina paired-end multiplexed sequencing platforms (Agilent Technologies), according to the manufacturer's sample preparation protocol. Briefly, 225 ng of genomic DNA was digested with restriction enzymes. The hybridization was performed for 3 hours at 54 °C, and the circularized target DNA-HaloPlex probe hybrids, containing biotin, were captured on streptavidin beads (HaloPlex Magnetics Beads, Agilent Technologies). The DNA with adaptor-modified ends was PCR amplified (number of cycles depended on the lot, Herculase II fusion DNA polymerase, Agilent). The amplified target DNA was purified using AMPure XP beads (Beckman Coulter Genomics, GENEWIZ, New Jersey, USA). All of the DNA samples were individually indexed. Amplification of the libraries was performed on a GeneAmp® PCR System 9700 thermocycler (Applied Biosystems). The restriction digestion and amplicon library quantities were quality evaluated using a Bioanalyzer High Sensitivity DNA Assay kit in an Agilent 2100 Bioanalyzer (Agilent Technologies) and quantified using a Qubit 2.0 Fluorometer (Invitrogen).

*TruSight One Panel (TSO).* TruSight One Sequencing Panel (Illumina, San Diego, CA) targeted 4,813 genes associated with a clinical phenotype. Libraries were generated using the TruSight One™ Sequencing Panel kit with the TruSight One™ Sequencing Panel (Illumina), according to the manufacturer's sample preparation protocol. Briefly, 50 ng of each DNA sample was enzymatically fragmented and adapter sequences were added to the ends. The fragmented DNA was purified, and barcodes and common adapters required for cluster generation and sequencing were PCR-added. After cleanup, 500 ng of each of the 12 DNA libraries was pooled. Then the libraries were hybridized twice to specific capture probes; the unhybridized material was washed away, and the captured fragments were amplified using PCR followed by purification. The enriched libraries were quantified using a Qubit 2.0 Fluorometer, and their quality was evaluated using a Bioanalyzer 2100 and the High Sensitivity DNA Kit (Agilent Technologies). Libraries were diluted and pooled to obtain the final sequencing equimolar pool.

*Whole Exome Sequencing (WES).* For whole exomes from 2011, the sample preparation for capturing approximately 44 Mb of selected human genome regions was performed according to the NimbleGen SeqCap EZ Exome Library SR protocol, v2.2, for Illumina paired-end sample libraries with modifications included in draft v1.4 (February 2011). In brief, 1.0 µg of genomic DNA was sheared on a Covaris™ E210 instrument (Covaris). The fragment size (150–400 bp) and the quantity were confirmed with the Agilent 2100 Bioanalyzer 1000 chip (Agilent). Fragmented DNA was prepared using an Illumina TruSeq DNA Sample Preparation Kit (Illumina) following the protocol described in the Illumina TruSeq DNA Sample Preparation Guide (revision A, November 2010) with the exception of the fragment size selection using an Agencourt SPRI XP Kit (Beckman Coulter). Instead of the standard Illumina PCR enrichment step, amplification via pre-capture LM-PCR (8 cycles) was performed with a SeqCap EZ Human Exome Kit v2.0 (Roche NimbleGen). One microgram of the amplified

library was hybridized to EZ probes at 47 °C for 72 hrs. After washing and recovery of the captured DNA, the library was amplified through post-capture LM-PCR (18 cycles). The final product was quality controlled on a Bioanalyzer DNA 1000 chip, and the success of the enrichment was measured with a qPCR SYBR Green assay on a LightCycler® 480 Instrument (Roche), evaluating one genomic locus with pre- and post-captured material.

For whole exomes from 2014, the NimbleGen SeqCap EZ v3.0 system for exome enrichment was used and pre-capture multiplexing was applied following the manufacturer's protocol version 4.2. Briefly, 1 µg of genomic DNA was fragmented with Covaris™ E210 and used for ligation of the adapters containing Illumina specific indexes with a KAPA Library Preparation kit (Kapa Biosystems). Adapter ligation DNA fragments were enriched by 7 cycles of pre-capture PCR using KAPA HiFi HotStart ReadyMix (2×) (Kapa Biosystems) and analyzed on an Agilent 2100 Bioanalyzer with the DNA 1000 assay. Five libraries were pooled with a combined mass of 1250 ng for the baits hybridization step (47 °C; 68 hrs). After washing (47 °C), the multiplexed captured library was recovered with capture beads and amplified with 14 cycles of post-capture PCR using KAPA HiFi HotStart ReadyMix (2×). The size, concentration and quality of the captured library were determined using an Agilent DNA 1000 chip. The success of the enrichment was measured using a qPCR SYBR Green assay on a Roche LightCycler® 480 Instrument evaluating one genomic locus with pre- and post-captured material.

**NGS: Run.** HPC and TSO libraries were sequenced on an Illumina Miseq instrument and NextSeq 500, respectively, following the manufacturer's protocol, with a paired end run. Each WES (2011 and 2014) library was sequenced on an Illumina HiSeq 2000 instrument in a fraction of a sequencing lane following the manufacturer's protocol, with a paired end run of 2×101 bp. Image analysis, base calling and quality scoring of the run were processed using the manufacturer's software Real Time Analysis (RTA 1.10.36 and RTA 1.13.48) and followed by generation of FASTQ sequence files by CASAVA.

**NGS: Data Analysis.** HPC and TSO data analyses were performed at Genycell Biotech S.L. (Madrid, Spain). Briefly, both panel sequencing reads were trimmed from the 3' end up to the first base with a Phred quality >9 and were mapped to the Human genome reference v37 with decoy sequences [hs37d5] (Broad, ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2\_reference\_assembly\_sequence/hs37d5.fa.gz) using the BWA-MEM version 0.7.5a. (bio-bwa.sourceforge.net/). To calculate read statistics, we used Prinseq-lite version 0.20.3 lite (prinseq.sourceforge.net/) and Picard Calculate HS Metrics version 1.119 (<http://picard.sourceforge.net>) to calculate metrics. Alignment (.bam) files containing only properly paired and uniquely mapped reads were processed with Picard Mark Duplicates version 1.119 to remove duplicates, and local realignment was performed with Freebayes version 9.9.13 (<https://github.com/ekg/freebayes>)<sup>43</sup>. SAMtools version 0.1.19 (<http://samtools.sourceforge.net/>)<sup>44</sup> was used on the processed BAM files to call single nucleotide variants (SNVs) and small insertion deletions (INDELs). Functional annotations from Ensembl release 75 (GRCh37.75 database)<sup>45</sup> were added to the resulting.vcf files using the Genome Analysis Tool Kit (GATK) version 2.4 (<https://www.broadinstitute.org/gatk/>)<sup>46</sup> to annotate variants with dbSNP version 137 and Ensembl Variant Effect Predictor version 72 (<http://www.ensembl.org/info/docs/tools/vep/index.html>) to annotate the variants.

For each WES (2011 and 2014), sequencing reads were trimmed from the 3' end up to the first base with a Phred quality >9 and were mapped to the Human genome reference v37 with decoy sequences [hs37d5] using the GEM toolkit<sup>47</sup>. Alignment (.bam) files containing only properly paired and uniquely mapped reads were processed with Picard tools version 1.110 (<http://picard.sourceforge.net>) to remove duplicates, and local realignment was performed with the Genome Analysis Tool Kit (GATK) version 3.1 (<https://www.broadinstitute.org/gatk/>)<sup>46</sup>. For alignments and coverage metrics, it has been remapped with BWA-MEM version 0.7.5a<sup>48</sup>. To call single nucleotide variants (SNVs) and small insertion deletions (INDELs) on the processed alignment (.bam) file, we used SAMtools version 0.1.19 (<http://samtools.sourceforge.net/>)<sup>44</sup>. Functional annotations from Ensembl release 75 were added to the resulting VCF using snpEff (<http://snpeff.sourceforge.net/>)<sup>49</sup>. SnpSift (<http://snpeff.sourceforge.net/SnpSift.html>)<sup>50</sup> was used to add information from dbSNP version 137, population frequencies from 1000 Genomes and the Exome Variant Server<sup>51</sup>, the NHLBI Exome Sequencing Project (<http://evs.gs.washington.edu/EVS/>) and a variety of conservation and deleteriousness predictions included in dbNSFP version 2.5 (<http://sites.google.com/site/jpopgen/dbNSFP>)<sup>52</sup>.

**Sanger validation.** The identified variants and familial segregation studies were validated by SS. The specific primers were designed online by Primer3 version 0.4.0 (<http://bioinfo.ut.ee/primer3-0.4.0/primer3/>). The PCR products were sequenced using a Big-Dye® Terminator version 3.1 Cycle Sequencing Kit in an Applied Biosystems 3730/DNA Analyzer (Applied BioSystems, Waltham, Massachusetts, USA). The raw data were analyzed with Chromas trace viewer (<http://technelysium.com.au/wp/chromas/>). The primers used for SS are shown in Supplementary Table S4.

## References

1. Rett, A. On a unusual brain atrophy syndrome in hyperammonemia in childhood. *Wien. Med. Wochenschr.* **116**, 723–726 (in German) (1966).
2. Hagberg, B., Aicardi, J., Dias, K. & Ramos, O. A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. *Ann Neurol* **14**, 471–479 (1983).
3. Laurvick, C. L. *et al.* Rett syndrome in Australia: a review of the epidemiology. *J Pediatr* **148**, 347–352 (2006).
4. Diagnostic criteria for Rett syndrome. The Rett Syndrome Diagnostic Criteria Work Group. *Ann Neurol* **23**, 425–428 (1988).
5. Hagberg, B., Hanefeld, F., Percy, A. & Skjeldal, O. An update on clinically applicable diagnostic criteria in Rett syndrome. Comments to Rett Syndrome Clinical Criteria Consensus Panel Satellite to European Paediatric Neurology Society Meeting, Baden Baden, Germany, 11 September 2001. *Eur J Paediatr Neurol* **6**, 293–297 (2002).
6. Neul, J. L. *et al.* Rett syndrome: revised diagnostic criteria and nomenclature. *Ann Neurol* **68**, 944–950 (2010).



7. Amir, R. E. *et al.* Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-CpG-binding protein 2. *Nat Genet* **23**, 185–188 (1999).
8. Miltenberger-Miltenyi, G. & Laccone, L. Mutations and Polymorphisms in the Human Methyl CpG-Binding Protein *MECP2*. *Mut Genet* **135**, 1343–1354 (2003).
9. Christodoulou, J., Grimm, A., Maher, T. & Bennetts, B. RettBASE: the IRSA *MECP2* variation database — a new mutation database in evolution. *Hum Mutat* **21**, 466–472 (2003).
10. Neul, J. L. *et al.* Specific mutations in methyl-CpG-binding protein 2 confer different severity in Rett syndrome. *Neurology* **70**, 1313–1321 (2008).
11. Chao, H. T. *et al.* Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* **468**, 263–269 (2010).
12. Neul, J. L. *et al.* Developmental delay in Rett syndrome: data from the natural history study. *J Neurodev Disord* **6**, 20 (2014).
13. Kalscheuer, V. M. *et al.* Mutations in the polyglutamine binding protein 1 gene cause X-linked mental retardation. *Nat Genet* **35**, 313–315 (2003).
14. Ariani, F. *et al.* FOXP1 is responsible for the congenital variant of Rett syndrome. *Am J Hum Genet* **83**, 89–93 (2008).
15. Armani, R. *et al.* Transcription factor 4 and myocyte enhancer factor 2C mutations are not common causes of Rett syndrome. *Am J Med Genet A* **158A**, 713–719 (2012).
16. Zhu, X. *et al.* Whole-exome sequencing in undiagnosed genetic disorders: interpreting 119 trios. *Genet Med* **17**, 774–781 (2015).
17. Sajjan, S. A. *et al.* Enrichment of mutations in chromatin regulators in people with Rett Syndrome lacking mutations in *MECP2*. *Genet Med* **19**, 13–19 (2017).
18. Sikkema-Raddatz, B. *et al.* Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. *Hum Mutat* **34**, 1035–1042 (2013).
19. Kurian, A. W. *et al.* Clinical evaluation of a multiple-gene sequencing panel for hereditary cancer risk assessment. *J Clin Oncol* **32**, 2001–2009 (2009).
20. Carvill, G. L. *et al.* GABRA1 and STXBP1: novel genetic causes of Dravet syndrome. *Neurology* **14**, 1245–1253 (2014).
21. Sweatt, J. D. Pitt-Hopkins Syndrome: intellectual disability due to loss of TCF4-regulated gene transcription. *Exp Mol Med* **45**, e21 (2013).
22. Zweier, M. *et al.* Mutations in MEF2C from the 5q14.3q15 microdeletion syndrome region are a frequent cause of severe mental retardation and diminish *MECP2* and *CDKL5* expression. *Hum Mutat* **6**, 722–733 (2010).
23. Bienvenu, T., Diebold, B., Chelly, J. & Isidor, B. Refining the phenotype associated with MEF2C point mutations. *Neurogenetics* **14**, 71–75 (2013).
24. Carvill, G. L. *et al.* Targeted resequencing in epileptic encephalopathies identifies de novo mutations in *CHD2* and *SYNGAP1*. *Nat Genet* **45**, 825–830 (2013).
25. Mignot, C. *et al.* Genetic and neurodevelopmental spectrum of *SYNGAP1*-associated intellectual disability and epilepsy. *J Med Genet* **53**, 511–522 (2016).
26. Fieremans, N. *et al.* Microdeletion of the escape genes *KDM5C* and *IQSEC2* in a girl with severe intellectual disability and autistic features. *Eur J Med Genet* **58**, 324–327 (2015).
27. Buiting, K. Prader-Willi syndrome and Angelman syndrome. *Am J Med Genet C Semin Med Genet* **154C**, 365–376 (2010).
28. Thiel, C. T. *et al.* A de novo 7.6 Mb tandem duplication of 14q32.2-qter associated with primordial short stature with neurosecretory growth hormone dysfunction, distinct facial anomalies and mild developmental delay. *Eur J Med Genet* **51**, 362–367 (2008).
29. Nava, C. *et al.* De novo mutations in *HCN1* cause early infantile epileptic encephalopathy. *Nat Genet* **46**, 640–645 (2014).
30. Lemke, J. R. *et al.* *GRIN2B* mutations in West syndrome and intellectual disability with focal epilepsy. *Ann Neurol* **75**, 147–154 (2014).
31. Carvill, G. L. *et al.* Mutations in the GABA Transporter *SLC6A1* Cause Epilepsy with Myoclonic-Atonic Seizures. *Am J Hum Genet* **96**, 808–815 (2015).
32. Brunklaus, A. & Zuberi, S. M. Dravet syndrome—from epileptic encephalopathy to channelopathy. *Epilepsia* **55**, 979–984 (2014).
33. Lucariello, M. *et al.* Whole exome sequencing of Rett syndrome-like patients reveals the mutational diversity of the clinical phenotype. *Hum Genet* **135**, 1343–1354 (2016).
34. Olson, H. E. Mutations in Epilepsy and Intellectual Disability Genes in Patients with Features of Rett Syndrome. *Am J Med Genet* **167A**, 2017–2025 (2015).
35. Gold, W. A. & Christodoulou, J. The utility of next-generation sequencing in gene discovery for mutation-negative patients with Rett syndrome. *Front Cell Neurosci* **9**, 266 (2015).
36. Lopes, F. *et al.* Identification of novel genetic causes of Rett syndrome-like phenotypes. *J Med Genet* **2016**, 53, 190–199 (2016).
37. Marx, V. Next-generation sequencing: The genome jigsaw. *Nature* **501**, 263–268 (2013).
38. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491–498 (2011).
39. Samorodnitsky, E. *et al.* Comparison of Custom Capture for Targeted Next-Generation DNA Sequencing. *J Mol Diagn* **17**, 64–75 (2015).
40. Sun, Y. *et al.* Next-Generation Diagnostics: Gene Panel, Exome, or Whole Genome? *Hum Mutat* **36**, 648–655 (2015).
41. Bebbington, A. *et al.* Investigating genotype-phenotype relationships in Rett syndrome using an international data set. *Neurology* **70**, 868–875 (2008).
42. Soto, D. *et al.* Rett-like severe encephalopathy caused by a de novo *GRIN2B* mutation is attenuated by D-serine dietary supplement. *Biol Psychiatry* **S0006-3223**, 31671–2, doi:https://doi.org/10.1016/j.biopsych.2017.05.028 (2017).
43. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907 [q-bio.GN]* (2012).
44. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (Oxford, England) **25**, 2078–2079 (2009).
45. Cunningham, F. *et al.* Ensembl 2015. *Nucleic Acids Res* **43**, D662–669 (2015).
46. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).
47. Marco-Sola, S., Sammeth, M., Guigó, R. & Ribeca, P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat Methods* **9**, 1185–1188 (2012).
48. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997 [q-bio.GN]* (2013).
49. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
50. Cingolani, P. *et al.* Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front genet* **3**, 35 (2012).
51. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308–311 (2001).
52. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum mutat* **34**, E2393–2402 (2013).

## Acknowledgements

We thank all patients and their families who contribute to this study. The work was supported by grants from the Spanish Ministry of Health (Instituto de Salud Carlos III/FEDER, PI15/01013); Crowdfunding program PRECIPITA, from the Spanish Ministry of Health (Fundación Española para la Ciencia y la Tecnología); Catalan Association for Rett Syndrome; Fondobio Rett and Mi Princesa Rett. The molecular analyses of eleven WES in 2014 were funded by the CNAG's 2013 call "300 exomes to elucidate rare diseases". AGC is supported by the ISCIII grant: FIS PI15/01082.

## Author Contributions

J.A. and M.P. conceived and supervised the study. S.V., P.P., N.B., E.G. and L.B., performed the experiments and collected the data. J.A., S.V., N.B., P.P., E.G., S.D., J.T. analyzed the results. J.A., M.O.C., A.G.C., M.P. provided samples and patients' clinical and genetic information. S.V., J.A., J.T., S.D. wrote the manuscript. All the authors reviewed the article critically for intellectual content.

## Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-11620-3](https://doi.org/10.1038/s41598-017-11620-3)

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017

## Consortia Rett Working Group

Francisco Javier Aguirre<sup>7</sup>, Montserrat Aleu<sup>8</sup>, Xènia Alonso<sup>4</sup>, Mercè Alsius<sup>9</sup>, Maria Inmaculada Amorós<sup>10</sup>, Guillermo Antiñolo<sup>11</sup>, Lourdes Aquino<sup>12</sup>, Carmen Arellano<sup>13</sup>, Gema Arriola<sup>14</sup>, Rosa Arteaga<sup>15</sup>, Neus Baena<sup>16</sup>, Montserrat Barcos<sup>17</sup>, Nuria Belzunces<sup>18</sup>, Susana Boronat<sup>19</sup>, Tomás Camacho<sup>20</sup>, Jaume Campistol<sup>4</sup>, Miguel del Campo<sup>19</sup>, Andrea Campo<sup>11</sup>, Ramon Cancho<sup>21</sup>, Ramon Candau<sup>11</sup>, Ignacio Canós<sup>22</sup>, María del Carmen Carrascosa<sup>23</sup>, Francisco Carratalá-Marco<sup>24</sup>, Jovani Casano<sup>25</sup>, Pedro Castro<sup>26</sup>, Ana Cobo<sup>27</sup>, Jaime Colomer<sup>4</sup>, David Conejo<sup>28</sup>, María José Corrales<sup>29</sup>, Rocío Cortés<sup>30</sup>, Gabriel Cruz<sup>31</sup>, Gábor Csányi<sup>32</sup>, María Teresa de Santos<sup>33</sup>, María de Toledo<sup>34</sup>, Mireia Del Toro<sup>19</sup>, Rosario Domingo<sup>35</sup>, Anna Duat<sup>36</sup>, Rosario Duque<sup>37</sup>, Ana María Esparza<sup>25</sup>, Rosa Fernández<sup>38</sup>, Maria Carme Fons<sup>4</sup>, Ana Fontalba<sup>15</sup>, Enrique Galán<sup>39</sup>, Pia Gallano<sup>32</sup>, María José Gamundi<sup>23</sup>, Pedro Luis García<sup>40</sup>, María del Mar García<sup>41</sup>, María García-Barcina<sup>42</sup>, María Jesús García-Catalan<sup>4</sup>, Sixto García-Miñaur<sup>43</sup>, Juan Jose García-Peñas<sup>36</sup>, María Teresa García-Silva<sup>44</sup>, Rosa Gassio<sup>4</sup>, Esther Geán<sup>4</sup>, Belén Gil<sup>38</sup>, Sarenur Gökben<sup>45</sup>, Luis Gonzalez<sup>36</sup>, Veronica Gonzalez<sup>4</sup>, Julieta Gonzalez<sup>4</sup>, Gloria González<sup>18</sup>, Encarna Guillén<sup>35</sup>, Miriam Guitart<sup>16</sup>, Montserrat Guitet<sup>46</sup>, Juan Manuel Gutierrez<sup>47</sup>, Eva Gutiérrez<sup>33</sup>, Jose Luis Herranz<sup>15</sup>, Gemma Iglesias<sup>48</sup>, Iva Karacic<sup>49</sup>, Carlos H. Lahoz<sup>50</sup>, José Ignacio Lao<sup>51</sup>, Pablo Lapunzina<sup>43</sup>, María Jesús Lautre-Ecenarro<sup>52</sup>, María Dolores Lluch<sup>53</sup>, Laura López<sup>36</sup>, Asunción López-Ariztegui<sup>54</sup>, Alfons Macaya<sup>19</sup>, Rosario Marín<sup>55</sup>, Charles M. Lourenço Marquez<sup>56</sup>, Elena Martín<sup>44</sup>, Beatriz Martínez<sup>38</sup>, Eduardo Martínez-Salcedo<sup>57</sup>, María José Mas<sup>58</sup>, Gonzalo Mateo<sup>14</sup>, Pilar Mendez<sup>39</sup>, Amparo Morant Jimenez<sup>59</sup>, Sira Moreno<sup>60</sup>, Fernando Mulas<sup>61</sup>, Juan Narbona<sup>62</sup>, Andrés Nascimento<sup>4</sup>, Manuel Nieto<sup>11</sup>, Tania Fabiola Nunes<sup>4</sup>, Núria Núñez<sup>19</sup>, María Obón<sup>9</sup>, Ignacio Onsurbe<sup>23</sup>, Carlos Ignacio Orteza<sup>4</sup>, Emilio Orts<sup>29</sup>, Francisco Martinez<sup>22</sup>, Rafael Parrilla<sup>63</sup>, Samuel Ignacio Pascual<sup>43</sup>, Ana Patiño<sup>64</sup>, María Pérez-Poyato<sup>4</sup>, Belén Pérez-Dueñas<sup>4</sup>, Pilar Póo<sup>4</sup>, Eliodoro Puche<sup>35</sup>, Feliciano Ramos<sup>65</sup>, Miquel Raspall<sup>19</sup>, Ana Roche<sup>4</sup>, Susana Roldan<sup>66</sup>, Jordi Rosell<sup>67</sup>, Cesar Ruiz<sup>68</sup>, María Luz Ruiz-Falcó<sup>36</sup>, Maria Eugenia Rusi<sup>4</sup>, Jordi Samarra<sup>69</sup>, Victoria San Antonio<sup>52</sup>, Ivan Sanchez<sup>4</sup>, Xavier Sanmartin<sup>4</sup>, Ana Sans<sup>4</sup>, Alfredo Santacana<sup>70</sup>, Sabine Scholl-Bürgi<sup>71</sup>, Nuria Serrano<sup>72</sup>, Mercedes Serrano<sup>4</sup>, Pilar Martin-Tamayo<sup>23</sup>, Adrián Tendero<sup>43</sup>, Jaime Torrents<sup>73</sup>, Diego Tortosa<sup>35</sup>, Emma Triviño<sup>74</sup>, Ledia Troncoso<sup>30</sup>, Eulalia Turrón<sup>32</sup>, Pilar Vázquez<sup>26</sup>, Carlos Vázquez<sup>70</sup>, Ramón Velázquez<sup>43</sup>, Clara Ventura<sup>19</sup>, Alfonso Verdú<sup>40</sup>, Anna Vernet<sup>4</sup>, M. Tomás Vila<sup>75</sup> & Cristina Villar<sup>4</sup>

<sup>7</sup>Hospital Torrecardenas, Almeria, Spain. <sup>8</sup>Consortio Hospital General Universitario de Valencia, Valencia, Spain. <sup>9</sup>Hospital Universitari Dr. Josep Trueta, Girona, Spain. <sup>10</sup>Hospital Punta Europa, Cadiz, Spain. <sup>11</sup>Hospital Universitario Virgen del Rocío, Sevilla, Spain. <sup>12</sup>Hospital de Mataró, Mataró, Spain. <sup>13</sup>Consorti Sanitari, Terrassa, Spain. <sup>14</sup>Hospital universitario de Guadalajara, Guadalajara, Spain. <sup>15</sup>Hospital Universitario Marqués de Valdecilla, Santander, Spain. <sup>16</sup>Hospital de Sabadell, Barcelona, Spain. <sup>17</sup>Hospital Universitario Reina Sofía, Córdoba, Spain. <sup>18</sup>Balagué Center, Barcelona, Spain. <sup>19</sup>Hospital de la Vall d'Hebrón, Barcelona, Spain. <sup>20</sup>Lema & Bandin Laboratorios, Vigo, Spain. <sup>21</sup>Hospital Universitario Río Hortega, Valladolid, Spain. <sup>22</sup>Hospital Universitario Doctor Peset, Valencia, Spain. <sup>23</sup>Hospital General Universitario de Albacete, Albacete, Spain. <sup>24</sup>Hospital Universitari San Juan, Alicante, Spain. <sup>25</sup>Hospital Universitario General de Castellón, Castellón de la Plana, Spain. <sup>26</sup>Hospital Gregorio Marañón, Madrid, Spain. <sup>27</sup>Hospital de Donostia, San Sebastian, Spain. <sup>28</sup>Complejo asistencial, Burgos, Spain. <sup>29</sup>Hospital General Mancha Centro, Ciudad Real, Spain. <sup>30</sup>Hospital San Borja Arriaran, Santiago, Chile. <sup>31</sup>Hospital Universitario de Valme, Sevilla, Spain. <sup>32</sup>Hospital de la Santa Creu i Sant Pau, Barcelona, Spain. <sup>33</sup>Hospital de Fuenlabrada, Madrid, Spain. <sup>34</sup>Hospital Universitario Severo Ochoa, Madrid, Spain. <sup>35</sup>Hospital Infantil de La Arrixaca, Murcia, Spain. <sup>36</sup>Hospital Infantil Universitario Niño Jesús, Madrid, Spain. <sup>37</sup>Hospital Universitario Nuestra Señora de la Candelaria, Santa Cruz de Tenerife, Spain. <sup>38</sup>Hospital Universitario de Getafe, Madrid, Spain. <sup>39</sup>Hospital Materno-Infantil de Badajoz, Badajoz, Spain. <sup>40</sup>Hospital Virgen de la Salud, Toledo, Spain. <sup>41</sup>Hospital Cormarcal de Figueres, Girona, Spain. <sup>42</sup>Hospital de Basurto, Bilbao, Spain. <sup>43</sup>Hospital La Paz, Madrid, Spain. <sup>44</sup>Hospital Universitario 12 de Octubre, Madrid, Spain. <sup>45</sup>Ege Ünüversütesü Tip Fakültesü Pedüatrü AD, Uzmur, Turkey. <sup>46</sup>Hospital General de Granollers, Barcelona, Spain. <sup>47</sup>Hospital Clínico Universitario de Valladolid, Valladolid, Spain. <sup>48</sup>Hospital Virgen de la Luz, Cuenca, Spain. <sup>49</sup>Clinical Hospital Center Zagreb, Zagreb, Croatia. <sup>50</sup>Hospital Central Asturias, Asturias, Spain. <sup>51</sup>Laboratorio Echevarne, Barcelona, Spain. <sup>52</sup>Hospital Clínico San Carlos, Madrid, Spain. <sup>53</sup>Hospital Universitario Virgen Macarena, Sevilla, Spain. <sup>54</sup>Hospital de Cruces, Bilbao, Spain. <sup>55</sup>Hospital Universitario Puerta del Mar, Cadiz, Spain. <sup>56</sup>Medical Genetics Service, Clinics Hospital of Ribeirão Preto, University of São Paulo, São Paulo, Brazil. <sup>57</sup>Hospital General Universitario de Alicante, Alicante, Spain. <sup>58</sup>Hospital Joan XXIII, Tarragona, Spain. <sup>59</sup>Centro privado, Valencia, Spain. <sup>60</sup>Hospital Virgen del Camino, Pamplona, Spain. <sup>61</sup>Instituto Valenciano de Neurociencias, Valencia, Spain. <sup>62</sup>Clínica Universitaria de Pamplona, Pamplona, Spain. <sup>63</sup>Complejo Hospitalario de Jaén, Jaén, Spain. <sup>64</sup>Hospital de Navarra, Navarra, Spain. <sup>65</sup>Hospital Clínico Universitario Lozano Blesa, Zaragoza, Spain. <sup>66</sup>Hospital Universitario Virgen de las Nieves, Granada, Spain. <sup>67</sup>Hospital Son Dureta, Palma de Mallorca, Spain. <sup>68</sup>Hospital Costa del Sol, Málaga, Spain. <sup>69</sup>Hospital General de Vic, Barcelona, Spain. <sup>70</sup>Complejo Hospitalario Universitario Insular, Las Palmas de Gran Canaria, Spain. <sup>71</sup>Medizinische Universität Innsbruck, Innsbruck, Austria. <sup>72</sup>Althaia, Manresa, Spain. <sup>73</sup>Reference Laboratory, Barcelona, Spain. <sup>74</sup>Catlab, Barcelona, Spain. <sup>75</sup>Hospital Francesc De Borja, Valencia, Spain.